Modeling Context Effects on Image Memorability

Zoya Bylinskii, Phillip Isola, Antonio Torralba, Aude Oliva
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Abstract: we use an information-theoretic framework to quantify context differences and image distinctiveness for predicting image memorability. We are able to quantify, using a large natural scene database, the observation that images that are unique or distinct with respect to their image context are better remembered.

The consistency of image memorability

- objective and quantifiable measure
- consistent across observers
- filter for visual data

$$\mathrm{HR(I)} = \frac{\mathrm{H(I)}}{\mathrm{H(I) + M(I)}} \times 100\%$$

$$\mathrm{FAR(I)} = \frac{\mathrm{FA(I)}}{\mathrm{FA(I) + CR(I)}} \times 100\%$$

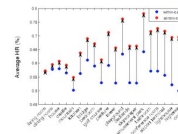## Crowd-sourced (AMT) memory (image recognition) games

ran 21 separate experiments: one per scene category → within-category experiment

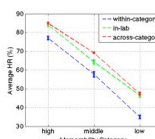an additional experiment, with images from all categories → across-category experiment

Memorability rank of images is consistent across participants and experiments
Spearman r = 0.69-0.86
(for each of 21 categories)

Memorability rank of categories is stable
Spearman r = 0.68
(across splits of images)

amusement park	64.2%
playground	63.3%
bridge	61.2%
pasture	59.2%
bedroom	58.9%
house	58.0%
dining room	57.8%
conference room	57.1%
bathroom	57.1%
living room	57.0%
castle	56.4%
kitchen	56.3%
airport terminal	55.6%
badlands	52.9%
golf course	52.9%
skyscraper	52.8%
tower	52.8%
lighthouse	52.1%
mountain	50.2%
highway	50.0%
cockpit	50.0%

Modeling image context

$$D(I; C) = -\log P_c(f_i)$$

context: set of images from which experimental sequence is sampled

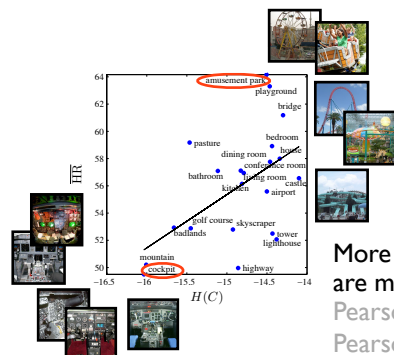$$P_c(f_i) = \frac{1}{\|C\|} \sum_{j \in C} K(f_i - f_j)$$

an image is **contextually distinct** if it is distinct with respect to the other images in its context
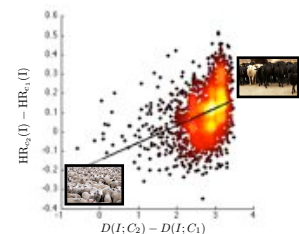
Contextually distinct images are more memorable
Pearson r(D₁,HR₁) = 0.26
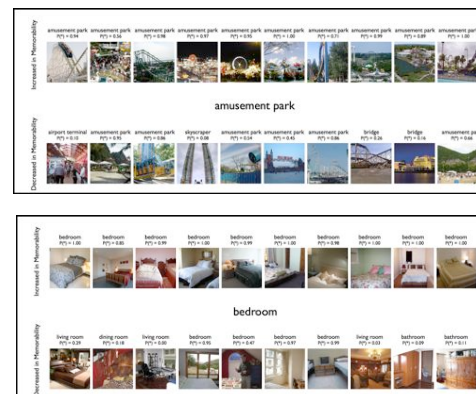Pearson r(D₂,HR₂) = 0.24
Pearson r(ΔD,ΔHR) = 0.35

Modeling is done with Places-CNN features which encode image semantics.

$$H(C) = \mathbb{E}_c[-\log P_c(f_i)]$$

More varied image contexts are more memorable overall
Pearson r(H,HR) = 0.53
Pearson r(ΔH,ΔHR) = 0.74

Images that look more like other scene categories, as measured by a scene classifier, are most affected by context

FIGRIM Dataset

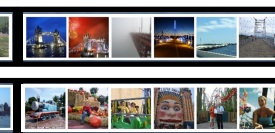21 SUN scene categories with more than 300 instances/category

	HR
amusement park	64.2%
pasture	59.2%
airport terminal	55.6%
mountain	50.2%
cockpit	49.5%
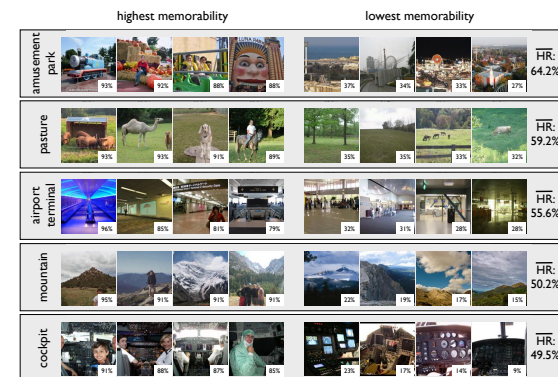
contains memorability scores for within-category and across-category experiments

Memorable within categories — more likely to look like images from other categories

Memorable across categories — more likely to be memorable across different contexts